

US Environmental Protection Agency Region 5

Electronic Data Processor Instructions

U.S. Environmental Protection Agency Region 5, 77 West Jackson Boulevard Chicago, IL 60604

June 2004

TABLE OF CONTENTS

1.0 INTRODUCTION	1
2.0 EDD DATA FILE CHECKS	1
3.0 INSTALLING EARTHSOFT'S EQUIS DATA PROCESSOR (EDP)	2
4.0 LOADING REFERENCE AND FORMAT FILES	10
5.0 UPDATING THE REFERENCE VALUE FILE	11
6.0 UPDATING THE FORMAT FILE	11
7.0 Using the EDP	12
7.1 Loading EDD Files	13
7.3 Correcting Errors	
7.4 Correcting Errors Directly Within EDP	18
7.5 Using Find and Replace	18
7.6 Saving Changes To the EDD File	
8.0 SUBMITTING EDD FILES TO EPA REGION 5	19

US ENVIRONMENTAL PROTECTION AGENCY – REGION 5 ELECTRONIC DATA PROCESSOR

1.0 INTRODUCTION

The Electronic Data Processor (EDP) has been made available to data providers in order to check their EDD files prior to submittal to EPA Region 5. The EDP is used to check EDD files to ensure they are formatted as described in the "Electronic Data Deliverable (EDD) Specification Manual" and the "EDD Manual for Historical Data". If the EDP detects errors, the errors will be identified and can be viewed directly within the EDP or via an error log. After the errors are corrected by the data provider, the EDP should be re-run to assure that no errors remain. The EDD files can then be sent to EPA Region 5.

The EDP is a product of EarthSoft, Inc. and replaces the two previous Region 5 EDD checker applications, the Electronic Laboratory Data Checker (ELDC) and the Electronic Field Data Checker (EFDC). The EDP is a single application that checks all EDD files and provides a much easier user interface for identifying and correcting errors.

This document's sole purpose is to assist EPA Region 5 Data Providers install and use the EDP in conjunction with submitting EDD data files to Region 5. Therefore, this document only provides information pertaining to the specific requirements of the Region 5 EDD specifications and is not intended to be a comprehensive EDP User's Manual. For more detailed discussion of the functionality and technical specifications of EDP, please refer to EarthSoft's web site at www.earthsoft.com.

2.0 EDD DATA FILE CHECKS

The EDP checks data for the following:

- Required Fields
- ➤ Field Lengths
- Data Types
- ➤ Valid Dates
- ➤ Reference Values
- Duplicate Rows
- ➤ Range Checking
- ➤ Record Parent-Child Relationships

Appendix A describes the checks performed by the EDP.

3.0 INSTALLING EARTHSOFT'S EQUIS DATA PROCESSOR (EDP)

Installation of EDP involves the following 4 steps:

- 1. Installing Microsoft's .NET Framework if not existing. The framework is a free download application available from Microsoft.
- 2. Downloading the EDP application from the EPA Region 5 website.
- 3. Installing and registering the EDP
- 4. Registering the EDP

Installing Microsoft's .NET Framework

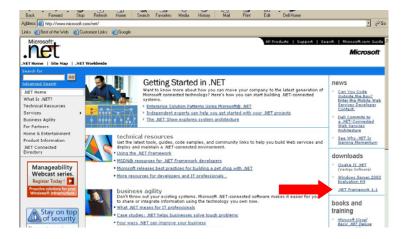
Earthsoft's EQuIS Data Processor (EDP) is a Visual Basic .NET application and requires Microsoft's .NET Framework. This application must be installed prior to the installation of the EDP.

The minimum system requirements for the .NET Framework are available on Microsoft's web page, and are as follows:

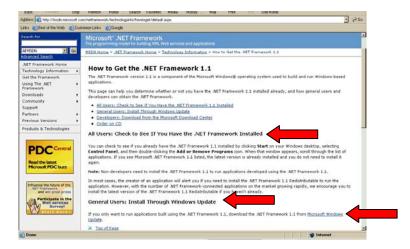
- Processor: 90-megahertz (MHz) Intel Pentium-class processor
- Windows Operating System: Windows 98, Me, NT4 SP6a, 2000, XP Home, XP Professional, and Server 2003
- Memory: 32 megabytes (MB) of RAM; 96 MB recommended
- Hard Disk: 150 MB of hard disk space required
- Display: 800 x 600 or higher-resolution display, at least 256 colors
- Input Device: Microsoft mouse or compatible pointing device
- Other: Microsoft Internet Explorer 5.01 or later Microsoft Data Access Components (MDAC) 2.6 or later

It should be noted that web pages illustrated in this document may vary slightly from the most current versions. In addition, this is not intended to be a definitive guide to installing any Microsoft utilities. If there any questions, you should consult Microsoft's web site at http://www.microsoft.com.

Installation of the .NET Framework is best done through the Microsoft Windows Update utility. This can be accessed from Microsoft's .net page: http://www.microsoft.com/net. On this page, a link, .NET Framework 1.1, to the .NET download is available under downloads.



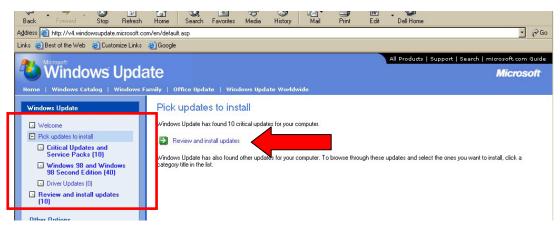
Clicking on the download link, .NET Framework 1.1, brings up the following page:



Follow the instructions under "All Users: Check to See If You Have .NET Framework Installed" to check if the .NET Framework is already installed on your computer. If it is, proceed to the next section of this document, "Downloading the EDP". If .NET is not installed, click on "Microsoft Windows Update" under the "General Users: Install Through Windows Upgrade" section. This should bring up the web page shown below.



Click 'Scan for Updates'. This will get the latest updates available for your computer's operating system, software, and hardware. There are 3 types of updates: 1) Critical Updates and Service Packs, 2) Windows and 3) Driver Updates. The result of a typical scan on a computer running Windows 98 is shown below.



In this case, 10 Critical Updates / Service Packs, 40 Windows and no Driver Updates were found. Microsoft considers the Critical Updates / Service Packs vital to peak performance of its operating systems. These should be installed before any other updates are attempted. The Update utility has automatically added these 10 to its "Review and Install Updates" selection. By clicking this link, the Critical Updates will be installed. It should be noted that in some instances, as is the case with an Internet Explorer update, the updates need to be installed separately from others. If this is the case, the utility will inform you as such. In some cases, you may have to restart the computer. You will then have to rerun the update utility. Instead of navigating back through the .net web pages, the utility can be directly accessed through Internet Explorer by selecting 'Windows Update' under the Tools menu. You can also access it directly through http://v4.windowsupdate.microsoft.com.

When all the Critical Updates / Service Packs are installed the Review and Install Updates link on the left will be grayed out. You are now ready to install the .NET Framework.



This .NET Framework 1.1 update is located under the Windows section. The name of this section will vary from computer to computer, but the standard is 'Windows' followed by the name of the operating system. In the above case, it is called 'Windows 98 and Windows 98 Second Edition'. Selecting this will bring up a list of 'Recommended Updates' as shown in the above screen.



Scroll through the list of Recommended Updates to locate "Microsoft .NET Framework version 1.1" and click the Add button. You will notice that the link on the left now shows 'Review and Install Updates (1). Click this link to perform the installation. When it is complete, you are ready to install EDP.

Downloading the EDP

The EDP installation application can be downloaded from EPA Region 5's website at www.epa.gov/region5/superfund/edman. The website will contain a link to the EDP. Click on the link and the EDP installation application will be downloaded. Save this installation file to a temporary directory.



Installing the EQuIS EDP

Open the directory where the EDP installation application was downloaded into, and double-click the file. The following instructions will step through the installation procedure. As the first screen suggests, you should have no other programs running during the installation.



Click the next button. The License Agreement screen will appear. Select 'I accept the license agreement' radio button and click next.



On the next screen, select the type of installation desired, Local or Network. For the installation folder, you may select the default or browse to another folder of your choosing. Click next.



Click the icon next to EPA Region 5 Format files and select 'Entire feature will be installed on local hard drive'.

Click the next button and you will be presented with the Editing and Auditing screen.



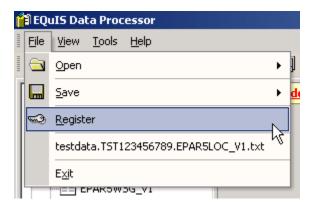
Selecting 'Yes' will allow you to make edits directly to the data in the EDD file via the EDP. Selecting 'No" will not permit any editing to the EDD file via the EDP. All edits to the EDD files must be done directly to the EDD text file after exiting the EDP. You also have the option of auditing all changes that are made to the data files. A log is created that includes the date and time, the user, the original value, and the new value. If Audit is selected, you will also have to select a directory to which the auditing files will be stored. Once the selection is made, click next. The EDP is now ready to be installed.



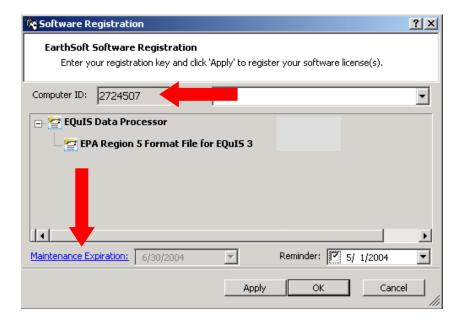
Click next to begin the installation. When the installation is done, you will be presented with a window that verifies that the EDP has been successfully installed. Click the Finish button to complete the installation.

Registering EDP

Once installed, EDP must be registered. The application will fully function for a limited time (1 to 2 days) after installation, however will resort to "Demonstration Mode" if not registered before the grace period. To register EDP:

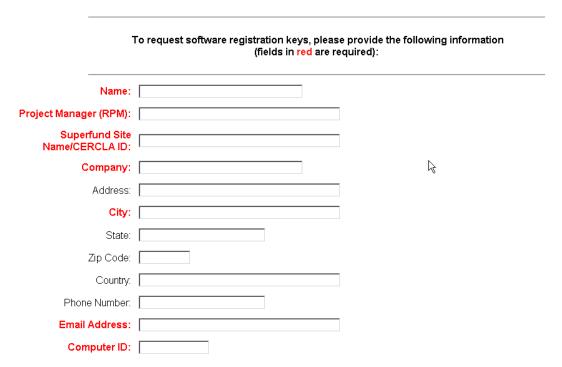


Select "Register" from the "File" pull down menu. The following registration screen will appear.

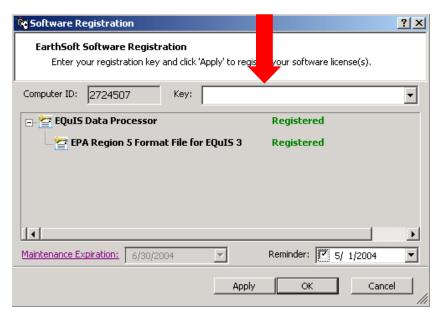


Record the Computer ID located in the upper left corner of the screen. Click on the "Maintenance Expiration" hyperlink. This will bring you to the EDP Registration Screen.

EPA Region 5 EQuIS Data Processor (EDP)



Fill in the requested information including your e-mail address and Computer ID. Click the "Submit" button located at the bottom of the form. A registration key will then be sent to the e-mail address provided in the EDP registration form.



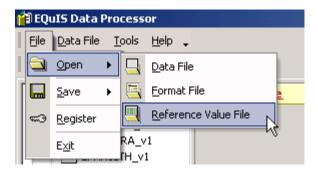
Open the above form by selecting "Register" from the "File" pull down menu. Enter the registration key in the "Key" field and select Apply. EDP is now registered.

4.0 LOADING REFERENCE AND FORMAT FILES

Before you can begin loading a data file to be checked in EDP, there are two important items that need to be in place: the Reference Value file and the Format File. These files determine how data files will be handled and affect the results of data quality checking. These files need only be loaded once after the installation of the EDP.

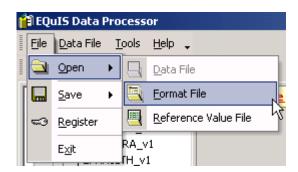
Note: After the initial install, EDP will be populated with a default Reference Value file and a default Format file. These default files are not the files to be used in the Region 5 EDP. The Region 5 Reference Value and Format files must be loaded to the EDP in order to properly check EDD files.

The Reference Value file must be loaded prior to the Format file. To load the Region 5 Reference Value file, select File>Open>Reference Value File from the menu bar.

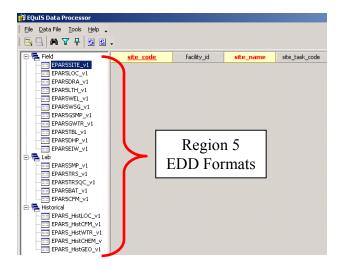


Browse to c:program files/earthsoft/equis5/interfaces/EPAR5/ and select the "epar5_refvals.xml" file and then click "Open". The Region 5 Reference File will then be loaded. After loading the Reference Value file, the Region 5 Format file needs to be loaded.

Select File>Open>Format File from the menu bar.



Browse to c:program files/earthsoft/equis5/interfaces/EPAR5/ and select the "EPAR5.xse" file and then click "Open". The Region 5 Format file will then be loaded. The Region 5 EDD formats will then be listed along the left side of the EDP.



5.0 UPDATING THE REFERENCE VALUE FILE

Periodically, EPA Region 5 will post an updated reference value file on the Region 5 EDMAN website www.epa.gov/region5/superfund/edman. Data Providers can update the reference values in their EDP application as follows:

- 1. Download the most recent reference value file from EPA Region 5's EDMAN website.
- 2. Replace the existing reference value file, located in the c:program files/earthsoft/equis5/interfaces/EPAR5/ folder, with the downloaded file.
- 3. Load the reference value file as described in the previous section.
- 4. Re-Load the existing format file as described in the previous section. Note: the existing format file must be reloaded in order for the EDP to read the updated reference value file.

6.0 UPDATING THE FORMAT FILE

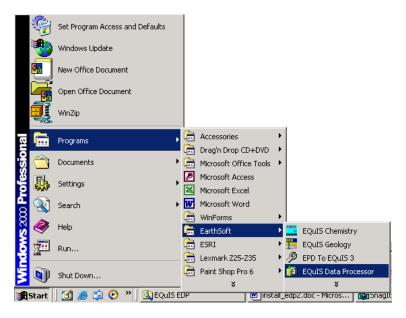
If Region 5 makes changes to the format of the existing EDD, the Data Provider will need to update their EDP application with a new format file. Data Providers can update the format file in their EDP application as follows:

- 1. Download the most recent format file from EPA Region 5's EDMAN website.
- 2. Replace the existing format file, located in the c:program files/earthsoft/equis5/interfaces/EPAR5/ folder, with the downloaded file.
- 3. Load the new format file as described in the previous section.

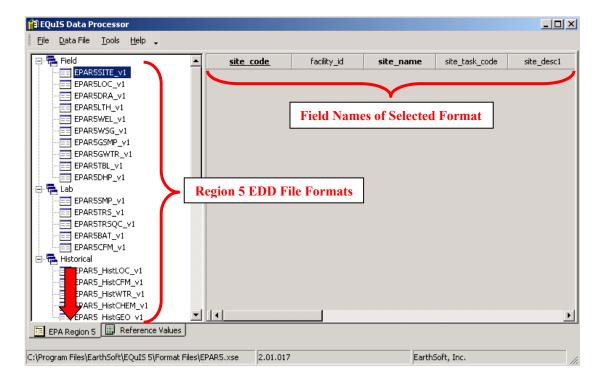
Note: Region 5 does not expect to make changes to the format in the near future, however, if changes are made, Region 5 will provide notification of the format changes to all data providers.

7.0 USING THE EDP

To start the application, select "EquIS Data Processor" from the Windows "Start" menu.



The following screen should appear showing the Region 5 file formats along the left side of the window and an empty table with the field names associated with the highlighted format along the top. If another screen appears, select the "EPA Region 5" tab located at the bottom left of the screen.



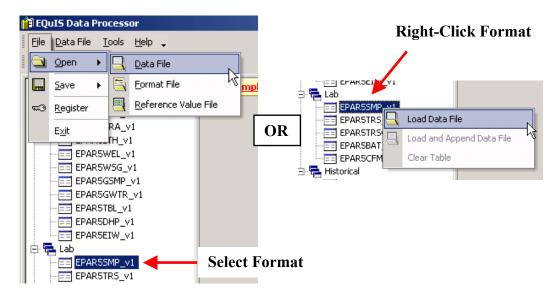
Each of the EDD file formats listed in the EDP corresponds to the EDD files described in the Region 5 EDD Specification Manual and the EDD Manual for Historical Data. In the following screen, the "EPAR5SITE_V1" format has been selected and its associated fields are displayed across the top.



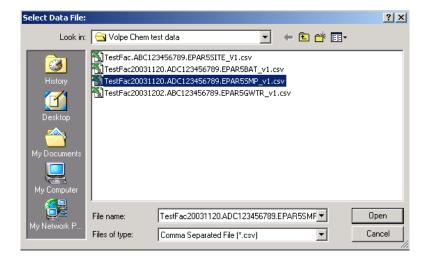
Fields with red text are "Required" fields and cannot be left blank, they must be populated with data. Information about each field is provided when the cursor is placed over the field name.

7.1 Loading EDD Files

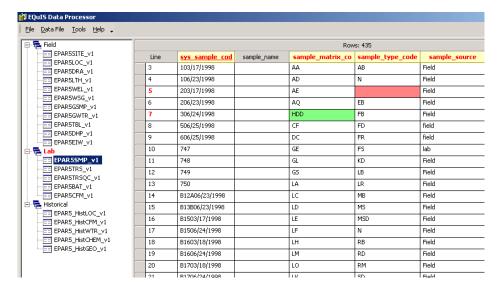
Files are checked by loading the individual EDD files into EDP.



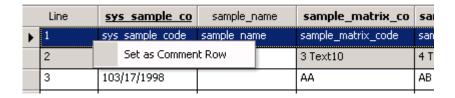
First, select the format type of the EDD file to be checked from the list. In the above example, an EPAR5SMP_v1 file is going to be checked, therefore, the EPAR5SMP_v1 format has been selected. Next, load the EDD data file by selecting the menu option File>Open>Data File or right-click on the format type and select "Load Data File"



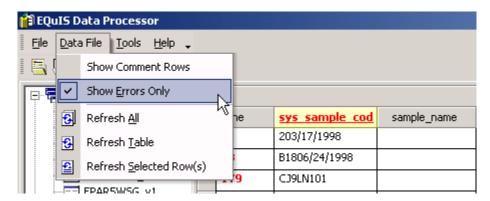
Browse to the EDD file and select "Open". The data file will load to the EDP and be checked during loading. Data will be displayed in the table and any detected errors will be shaded. Red line numbers indicate errors in the row.



In the example above, errors were detected in lines 5 and 7. <u>Note</u>: If the data file contains headers, EDP will identify the header rows as errors. To hide these rows, highlight the header row, right-click the row, and select "Set as Comment Row". To unhide the row, select "Show Comment Rows" from the "Data File" menu located at the top of the EDP window.



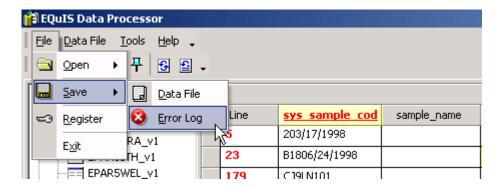
To view only the rows with errors, select Data File>Show Errors Only. To restore the view of all the records, deselect the Show Errors Only option.



To clear the data from EDP, right click the EDD Format and select "Clear Table". The EDD file will be cleared from the EDP viewer. Note: clearing the data from the EDP will not delete the EDP file, it only removes the file from the viewer.

7.2 Error Logs

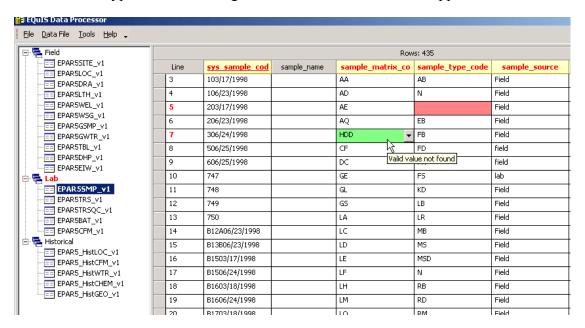
EDP produces an error log that can be saved as a file (html fomat). To save the error log, select File>Save>Error Log.



Browse to the desired folder and select "Save". The error log will then be saved in the selected folder.

7.3 Correcting Errors

As stated above, data are being checked for errors by the EDP as the data files are loading. The fields with errors will be shaded different colors depending on the type of error. The types of errors being checked for are described in Appendix A.

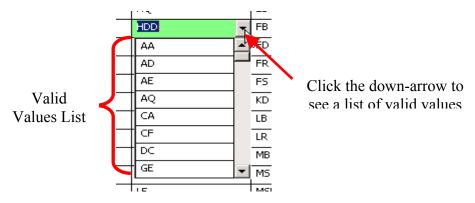


A description of the error is provided when the cursor is placed over the field. In the above example the sample_matrix_code value in line 7 is not a valid value. Two methods can be used to correct the error:

- 1. The user can exit the EDP and then open the EDD file using a text editor or spreadsheet application. Errors are then corrected, the file saved, and then loaded once more to the EDP to ensure no further errors.
- 2. The user can correct errors directly within the EDP application by clicking in the field and then changing the value. Note, this method can only be used if "Yes" or "Audit" was selected in the "Editing and Auditing" screen during EDP installation

7.4 Correcting Errors Directly Within EDP

To correct errors directly within EDP, click in the error field and type the correct value. If the field is restricted to a list of valid values, a list of the values will be provided from a drop down by clicking on the down arrow located on the left side of the field.



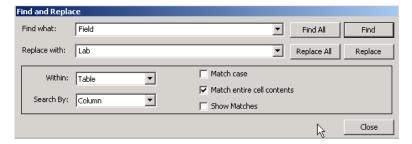
Select the appropriate value from the list or type the value directly into the field. Once a valid value is selected and the cursor is moved out of the field (i.e. user clicks on another field) the shading signifying an error should disappear.

7.5 Using Find and Replace

The "Find and Replace" function allows searching the file for a specified value and then replacing it with another value. This function is useful when there are a number of similar values that need to be changed. The "Find and Replace" function is activated by selecting Tools>Find or by clicking the "Find" icon located in the menu bar.



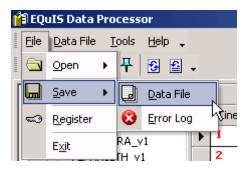
The "Find and Replace" dialog will appear.



Type the value to be replaced in the "Find What" field and the new value in the "Replace With" field and select "Find" to view fields with the value and "Replace" to replace the original value with the new value.

7.6 Saving Changes To the EDD File

To save the changes made to the EDD data file, select File>Save>Data File.



Browse to the folder where the data file is to be saved. Any changes made to the EDD will now be saved to the file.

8.0 SUBMITTING EDD FILES TO EPA REGION 5

After the EDD files have been checked by the EDP and no errors are identified the files are ready for submittal to EPA Region 5. Please follow the procedures for submitting EDD files described in Region 5's "Electronic Data Deliverable (EDD) Specification Manual" or "EDD Manual for Historical Remedy Implementation Data" located at EPA Region 5's website, http://www.epa.gov/region5superfund/edman.

APPENDIX A

This appendix describes the errors identified by the EDP.

1. Valid Value Not Found and Value Not Found on List

The value in the field does not match the values listed in the Valid Value Appendix of the EDD Specification Manual. Use the field description in the EDD Specification Manual for details of values that are required. If an appropriate value does not exist, contact EPA Region 5 to request that a new valid value be added to the valid value table.

2. Value Exceeds Field Length

The number of characters of the value entered in the field exceeds the maximum allowed number of characters. See the EDD Specification Manual for the maximum field lengths.

3. Missing Required Field

The field must be populated with a value. The field cannot be left null (i.e., blank).

4. Invalid Data Type

The value is not the appropriate data type. Each field has a specific data type that must be used, such as text, date, time, or number. If the appropriate data type for a field is 'data', then the value must be a date in MM/DD/YYYY format. See the Data Type description in the EDD Specification Manual for the appropriate data type.

5. Out of Range

The value is not within the allowable range of values. The following fields have values that must be within an allowable range:

Field	EDD File Format	Allowable Value
		Range
dilution_factor	EPAR5TRS	greater than 1.0
	EPAR5TRSQC	
final_volume	EPAR5TRS	greater than 0.0
	EPAR5TRSQC	
result_value	EPAR5TRS	greater than 0.0
	EPAR5TRSQC	
reporting_detection_limit	EPAR5TRS	greater than 0.0
	EPAR5TRSQC	

Field	EDD File Format	Allowable Value
		Range
qc_original_conc	EPAR5TRSQC	greater than 0.0
qc_spike_added	EPAR5TRSQC	greater than 0.0
qc_spike_measured	EPAR5TRSQC	greater than 0.0
qc_spike_recovery	EPAR5TRSQC	greater than 0.0
qc_dup_original_conc	EPAR5TRSQC	greater than 0.0
qc_dup_spike_added	EPAR5TRSQC	greater than 0.0
qc_dup_spike_measured	EPAR5TRSQC	greater than 0.0
qc_dup_spike_recovery	EPAR5TRSQC	greater than 0.0

6. Duplicate Row

Two or more records have the same values in the primary key fields. The primary key fields are the fields that make each record in the file unique. No two records can have the same values in the primary keys. For example, the EPAR5LOC file has the sys_loc_code and sys_well_code fields as the primary keys. Two records that both have "MW-01" and "Well-1" in the sys_loc_code and sys_well_code fields, respectively, would be considered duplicate records. To make each record unique, one record would have to change either the sys_loc_code to something other than "MW-01" and/or change the sys_well_code to something other than "Well-1". Refer to Section 2.6 "Data Integrity Rules" of the Region 5 EDD Specification Manual for further discussion of duplicate records.

7. Orphan Row

The record is missing a required parent record. Records that depend on information (i.e., child records) from another record (parent record) must reference the parent record and the parent record must exist in the corresponding file. For example, each row in the EPAR5TRS file must include a sys_sample_code that corresponds to a sys_sample_code reported in the EPAR5SMP file. If a record in the EPAR5TRS file has a sys_sample_code of GWSMP-006 then a record must also be included in the EAPR5SMP file with a sys_sample_code of GWSMP-006. If a record in the EPAR5TRS file has a sys_sample_code that is not included in the EPAR5SMP file, an "orphan record" error will be identified. For a further discussion of child/parent records see section 2.6 "Row Integrity" in the Region 5 EDD Specification Manual.

Note: When checking the EPAR5BAT file, the *EPAR5TRSQC* file must be checked through the EDP prior to checking the EPAR5BAT file, otherwise all records in the EPAR5BAT file will be identified as "Orphan Row" Errors. The EPAR5BAT file is only submitted if a EPAR5TRSQC file is submitted.

8. Repotable_result Cannot be "Yes" When lab_qualifiers = E, G, P, or R

Identifies records as errors that have a laboratory qualifier that signifies the result should not be reportable, yet the record has "Yes" reported in its reportable_result (EPAR5TRS and EPAR5TRSQC, Pos #36) field. If a record has a lab_qualifier (EPAR5TRS and EPAR5TRSQC, Pos #38) value of "E", "G", "P" or "R" they are considered as not reportable and the reportable_result (Pos #36) field must be populated with "No".

9. Result_value is Required when detect_flag = Y

Identifies records that have the detect_flag (EPAR5TRS and EPAR5TRSQC, Pos #37) value of "Y" yet there is no value reported in the result_value field (EPAR5TRS and EPAR5TRSQC, Pos #33). This error applies only to records of target analytes (TRG) and tentatively identified compounds (TIC). If a record has a value of "TRG" or "TIC" in the result_type_code (EPAR5TRS and EPAR5TRSQC, Pos #35) and the detect_flag (EPAR5TRS and EPAR5TRSQC, Pos #37) value of "Y", the result_value field (EPAR5TRS and EPAR5TRSQC, Pos #33) must be populated with the test result value (i.e., it cannot be left blank).

10. Reporting detection limit Cannot be Null when detect flag = N

Identifies records with detect_flag (EPAR5TRS and EPAR5TRSQC, Pos #37) value of "N" and the reporting_detection_limit (EPAR5TRS and EPAR5TRSQC, Pos #20) field is null. All records that have a value of "N" in the detect_flag field must have the reporting_detection_limit field populated with the appropriate detection limit value (i.e., it cannot be left null).

11. Parent_sample_code is Required Where sample_type_code = BD, FD, FR, FS, MS, SD, MSD

Identifies records that have a sample_type_code (EPAR5SMP, Pos #4) of 'BD', 'FD', 'FR', 'FS', 'LR', 'MS', 'SD', 'MSD' but are missing the appropriate parent_sample_code (EPAR5SMP, Pos #6). The above sample_type_codes signify duplicates, and the sample identifier (i.e., sys_sample_code) of the original sample from which the duplicate was derived must be populated in the parent_sample_code (EPARSMP, Pos #6) field. The parent_sample_code value must match the sys_sample_code of the original sample and the original sample must also be reported as a separate record in the EPAR5SMP file (i.e., there should be a record for the original sample and a separate record for the duplicate sample).

12. Dates Not in Sequence

Errors will be identified if certain dates precede other dates, such as the analysis date preceding the sampling date. The following fields are checked for the proper date sequence:

Field	EDD Format File	Proper Sequence
Calibration_date	EPAR5CFM_V1	calibration_date must
Measurement_date		precede the
		measurement_date.
Sample_date	EPAR5SMP_V1	sample_date must
Analysis_date	EPAR5TRS_V1	precede the
	EPAR5TRSQC_V1	analysis_date.
Sample_date	EPAR5SMP_V1	sample_date must
Prep_date	EPAR5TRS_V1	precede prep_date
	EPAR5TRSQC_V1	
Sample_date	EPAR5SMP_V1	sample_date must
leachate_date	EPAR5TRS_V1	precede leachate_date.
	EPAR5TRSQC_V1	

13. qc_spike_measured Less Than qc_original_conc and qc dup spike measured Less Than qc dup original conc

The qc_original_conc and the qc_dup_original_conc must be greater than qc_spike_measured and qc_dup_spike_measured, respectively. This applies to the EPAR5TRSQC format.

14. sys_loc_code is Required when sample_type_code = "N"

A location identifier, i.e., sys_loc_code, must be provided for all samples that are normal environmental samples. Therefore, all records in the EPAR5SMP_V1 file that have a sample_type_code (Pos #3) of "N" must also have the sys_loc_code field (Pos #10) populated (i.e., this field cannot be left blank).